(PSI)

POSTSCRIPT INPUT FILE FOR NEWSPAPER PAGE

EXTRACT TEXT

EXTRACT TEXT POSITION

EXTRACT FONT INFORMATION

EXTRACT IMAGE POSITIONS

EXTRACT BITMAP OF PAGE

EXTRACT PAGE REFER DATA

EXTRACT URL/EMAIL DATA

EXTRACT PAGE STORY LOCATIONS

EXTRACT IMAGE AD LOCATIONS

EXTRACT BOOKMARK DATA

CURRENT PAGE INFORMATION DATABASE

(CPDB)

FIG. 1

(CPDB)

CURRENT PAGE
INFORMATION
DATABASE

(PPDB)

PREDEFINED
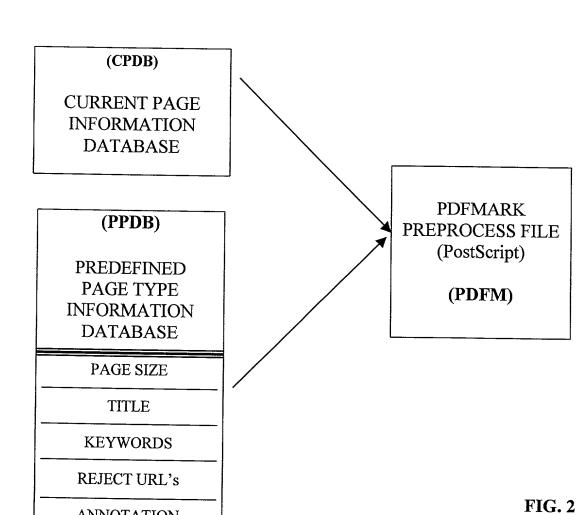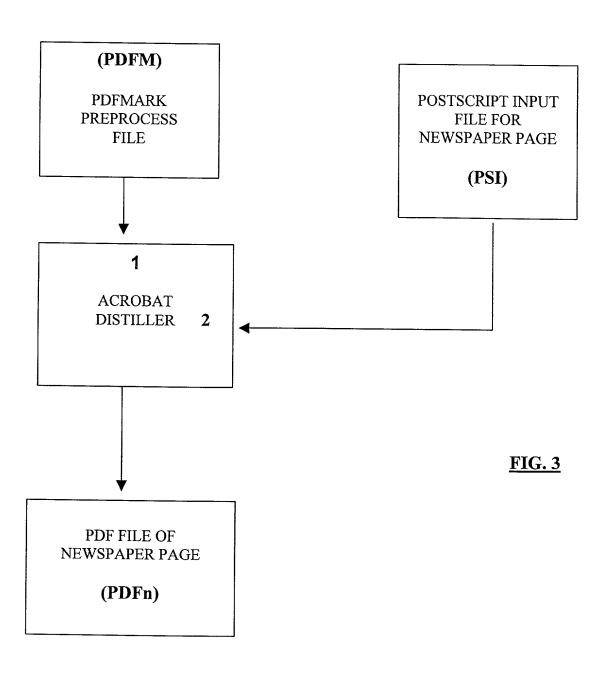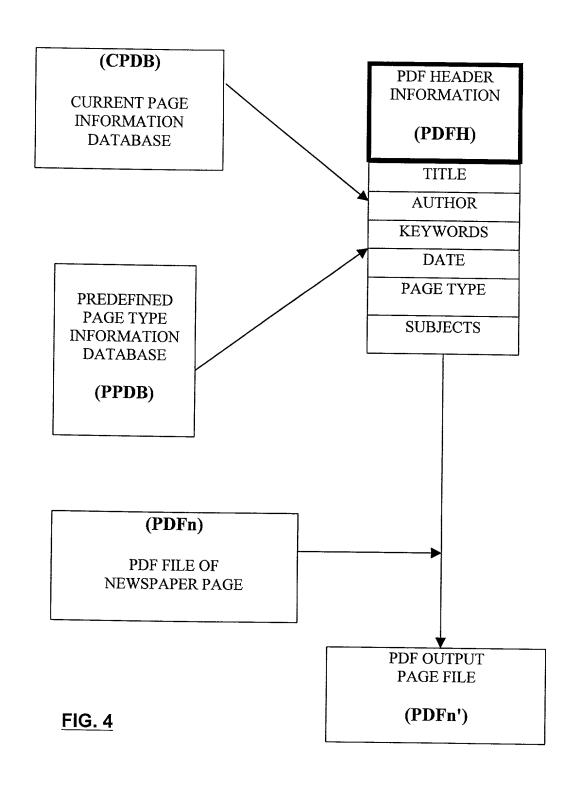PAGE TYPE
INFORMATION
DATABASE

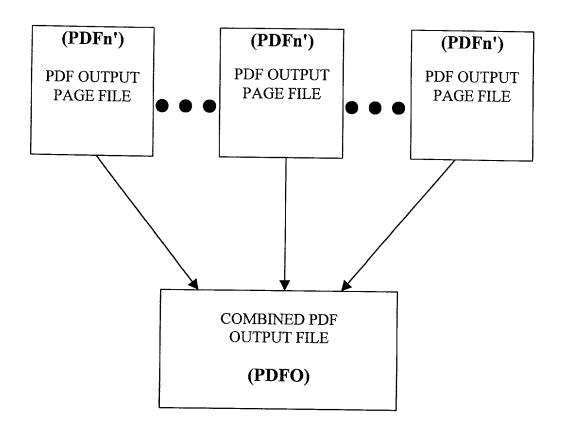| PAGE SIZE |
| TITLE |
| KEYWORDS |
| REJECT URL's |
| ANNOTATION INFORMATION |
| PREDEFINED PAGE REFERS |

PDFMARK
PREPROCESS FILE
(PostScript)

(PDFM)

FIG. 2

```
(PDFM)

PDFMARK
PREPROCESS
FILE
```

```
POSTSCRIPT INPUT
FILE FOR
NEWSPAPER PAGE

(PSI)
```

```
1

ACROBAT
DISTILLER     2
```

```
PDF FILE OF
NEWSPAPER PAGE

(PDFn)
```

FIG. 3

(CPDB) CURRENT PAGE INFORMATION DATABASE	**PDF HEADER INFORMATION** **(PDFH)**
	TITLE
	AUTHOR
	KEYWORDS
	DATE
	PAGE TYPE
	SUBJECTS

PREDEFINED PAGE TYPE INFORMATION DATABASE

(PPDB)

(PDFn)

PDF FILE OF NEWSPAPER PAGE

PDF OUTPUT PAGE FILE

(PDFn')

FIG. 4

(PDFn')

PDF OUTPUT
PAGE FILE

● ● ●

(PDFn')

PDF OUTPUT
PAGE FILE

● ● ●

(PDFn')

PDF OUTPUT
PAGE FILE

COMBINED PDF
OUTPUT FILE

(PDFO)

FIG. 5